

DOCUMENT RESUME

ED 352 393

TM 019 297

TITLE The Future of the National Assessment of Educational Progress. Discussion Paper.

INSTITUTION National Assessment Governing Board, Washington, DC.

PUB DATE Aug 92

NOTE 35p.; Prepared by the Ad Hoc Committee on the Future of NAEP.

PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Academic Achievement; *Academic Standards; *Educational Assessment; Educational History; Educational Policy; Elementary Secondary Education; *Futures (of Society); International Studies; *National Programs; *Policy Formation; Program Development; Public Opinion; Research Methodology; Student Evaluation; *Testing Programs; Test Results; Test Use

IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

The 1990s show distinct changes from the 1960s in terms of the views held by education practitioners, policymakers, and the public concerning the National Assessment of Educational Progress (NAEP). Developments such as the national education goals, the prospect of voluntary national education standards, and the possibility of a national system of assessments raise several issues for the NAEP. These issues include: (1) the role and purpose of the NAEP; (2) the alignment with national education standards; (3) assessment frameworks; (4) the role of the NAEP in relation to organizations that may be established to review or certify national standards and a system of assessments; (5) NAEP achievement levels; (6) an international component to the NAEP; (7) the NAEP as an anchor for linking state and local assessment systems with national and international results; (8) removing the prohibition against using NAEP results at the district or school level; and (9) annual assessment and reporting. Policy questions associated with each of these issues are discussed. (SLD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *



National Assessment Governing Board

National Assessment of Educational Progress

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

Discussion Paper on

The Future of

The National Assessment of Educational Progress

Prepared by
The Ad Hoc Committee on
The Future of NAEP
August, 1992

BEST COPY AVAILABLE

800 North Capitol Street, N.W.
Suite 825
Washington, D.C. 20002
(202) 357-6938

ED352393

T49019297

Table of Contents

	Page
1. Acknowledgements	ii
2. Executive Summary	iii
3. Preface	vi
4. Background	1
5. Role and Purpose of the National Assessment of Educational Progress	4
6. Alignment with National Standards	9
7. Assessment Frameworks	11
8. Relation to Standards and Assessment Review and Certification	12
9. Achievement Levels	15
10. An International Component	17
11. The National Assessment as a Comparability Anchor	19
12. Removing the Prohibition against Using National Assessment Results at the District or School Level	21
13. Annual Assessments and Reporting	23
14. Conclusion	25
15. Epilogue	26

ACKNOWLEDGEMENTS

Grateful appreciation is expressed for the thoughtful, incisive comments provided on an early draft of this discussion paper by Francie Alexander, Gordon Ambach, Thomas Boysen, Christopher Cross, Paul H. Koehler, Bernard H. Martin, Wayne Martin, Ina Mullis, Eugene T. Paslov, Rodney Riffle, Lisa Ross, and Frank B. Womer. Of course, the content of this discussion paper is the sole responsibility of the National Assessment Governing Board.

EXECUTIVE SUMMARY

The National Assessment of Educational Progress, a congressionally authorized program, is this nation's only continuing, nationally representative indicator of education performance for students at grades 4, 8 and 12 and for young adults. Conceived in 1963 and first conducted in 1969, the National Assessment¹ has earned a reputation for integrity, innovation and quality. When it was created, virtually no data were available on the effectiveness of American education. The establishment of the National Assessment as a regular education indicator system represented a remarkable advance. However, concerns in the 1960s about the establishment of a national curriculum and the possible erosion of local control of education by the federal government influenced the design of the National Assessment.

In the 1990s, there are distinct changes from the 1960s views of the National Assessment by education practitioners, policymakers and the public. The most pronounced expressions of these changes include the establishment of national education goals, the prospect of voluntary national education standards, and the possibility of a national system of assessments. These developments raise a number of issues for the National Assessment. These issues, and related questions, are addressed in this report. They include the following:

1. Role and Purpose of the National Assessment

Should the National Assessment continue as designed in the 1960s as a national and regional monitor only?

Should the National Assessment regularly collect and report state-level data?

2. Alignment with National Education Standards

Should national content and performance standards determine the content of each assessment or should the National Assessment reflect these standards and the current (and evolving) instructional programs?

What formal mechanisms should the National Assessment institute in order to collaborate with organizations as they develop standards?

¹ The terms National Assessment of Educational Progress and National Assessment, and the acronym NAEP, are used interchangeably.

3. Assessment Frameworks

Should the National Assessment be designed in a way that leads instruction (as developed through a national consensus approach) or should it reflect representative instructional practice?

Should the National Assessment assess subjects in addition to reading, mathematics, science, writing, geography and U.S. history?

Should the approach to measuring trends be modified?

4. Role of the National Assessment in Relation to Organizations that May be Established to Review/Certify National Standards and a System of Assessments

Should the National Assessment, through its Governing Board and the National Center for Education Statistics, be treated the same as other assessment programs by any new entity created to review/certify standards and assessments?

Should there be a relationship specified in law between the National Assessment and any new entity created to review/certify standards and assessments? If so, what should that relationship be?

5. National Assessment Achievement Levels

Do achievement levels improve public understanding of the National Assessment?

Should National Assessment results continue to be reported using achievement levels?

Should other approaches to identifying appropriate achievement goals be considered? If so, which approaches should be considered?

6. An International Component to the National Assessment

Would it be useful to see how students in other nations perform on the National Assessment or on certain NAEP items as a reference point for understanding U.S. student performance?

Should the development of National Assessment frameworks regularly take into account the nature of curricula, instruction and expectations of foreign education systems?

Should the National Assessment achievement levels be developed taking into account student performance in other nations?

7. The National Assessment as an Anchor for Linking State and Local Assessment Systems with National and International Results

Should states and local districts be permitted to use the National Assessment as an anchor test for comparability purposes?

Should the federal government provide resources to permit research and development for such uses of the National Assessment?

Should there be federal oversight of such uses of the National Assessment to protect the integrity of the National Assessment, to avoid abuses and to assure that linking procedures are properly conducted?

8. Removing the Prohibition against Using National Assessment Results at the District or School Level

Should states and local districts be allowed to use the National Assessment for local assessment at the district and school levels if they wish to do so under National Assessment regulations and at their own option and cost?

9. Annual Assessment and Reporting

Should the National Assessment legislation be amended to permit annual assessments?

The National Assessment Governing Board hopes that this discussion paper will prompt consideration of these issues. The Governing Board invites readers of this paper to share their comments, perspectives and suggestions. The Governing Board will use comments received to help develop its positions on these issues. Initial positions on these issues will be prepared for discussion at the November 19-21, 1992 meeting of the National Assessment Governing Board.

PREFACE

The National Assessment Governing Board is charged by Congress to formulate policy for the National Assessment of Educational Progress. More specifically, the Board is assigned the responsibility for "taking appropriate actions needed to improve the form and use of the National Assessment."

Considering this responsibility, the recognition that the circumstances present at its inception have changed significantly, and the prospect of a national system of standards and assessments, the Governing Board wishes to begin a public dialogue on whether and how the National Assessment should evolve as a monitor of national education goals for student achievement.

The world, the needs of American students, and the policy context of American education have changed significantly since 1963, when the National Assessment of Educational Progress was conceived by U.S. Commissioner of Education, Francis Keppel.

Technology brings instantaneous contact with virtually every part of the globe; international markets and economic systems are increasingly competitive and interdependent; and the prospect of democratic governments and market economies replacing authoritarian governments and centrally controlled economies all pose new challenges for today's U.S. citizens. The education system must equip our society to meet these challenges by preparing students to perform to their full intellectual capacity, to compete effectively in the marketplace, and to participate thoughtfully in our democratic system of government.

The growing recognition of international challenges has been accompanied by disturbing indications that the American education system may not be producing sufficiently well-prepared students. The establishment of national education goals and the call for rigorous, voluntary national education standards are responses to these concerns which may bear significantly on the future role and purpose of the National Assessment.

While it has developed into a program of recognized integrity and credibility and has often led the way in assessment technology, the fundamental assumptions that undergird the National Assessment are largely unchanged since its inception. Given developments in the world over the last three decades and the increased expectations placed on the education system to prepare society to meet the Nation's challenges, it is appropriate to consider the implications of these changes for the National Assessment.

BACKGROUND

In 1963, concerned that virtually no data existed to describe the effectiveness of American education, U.S. Commissioner of Education, Francis Keppel, asked Ralph Tyler to prepare a paper "outlining procedures by which necessary information might be periodically collected to furnish a basis for public understanding of educational progress and problems." This set in motion a series of conferences and design efforts, supported by grants from the Carnegie Corporation, the Ford Foundation and the Fund for the Advancement of Education. In 1969, the National Assessment was created.

There was broad consensus in support of its creation. Many saw the value in having dependable, comprehensive information on the progress of education in this country. But there also were concerns. Some feared that National Assessment could lead to a national curriculum; they felt that it should be an unobtrusive measure of what the schools were teaching. Others were concerned about local autonomy; they did not want to see the federal role in education expanded in a way that would undercut state and district-level decision-making. Still others worried that a national test would provide a basis for accountability measures that were not universally welcomed. The National Assessment planners addressed these concerns.

First, the results were to be reported only for the nation and regions; **there would be no state, district, school, or individual student results.** In addition, several design decisions made for technical reasons also had the effect of addressing the concerns. For example, student samples were defined by age (nine, thirteen, seventeen and twenty-six to thirty-five year-olds); **thus, no grade-level results could be reported.** Further, matrix sampling was used to provide comprehensive content area coverage and reduce individual student burden, but this also meant that only aggregate results could be reported since no student would take a complete test. Finally, to keep the federal government at "arms length," the assessment would be carried out under a grant. From 1969 to 1983 the grant was awarded to the Education Commission of the States. Some commentators have called the original design of the National Assessment "brilliantly responsive to the political constraints of the time."²

² Messick, Samuel J., Beaton, Albert E., and Lord, Frederic M. (1983). National Assessment of Educational Progress Reconsidered: A New Design for a New Era. Princeton, NJ: Educational Testing Service, quoted in an unpublished paper by Albert E. Beaton.

During the 1980's, a shift began in the nation's concern about education and student assessment. The release of "A Nation at Risk" in 1983 brought national attention to the need for education reform. State reform efforts focused on more demanding curricula, more rigorous graduation requirements, heightened concern for teacher competency, and school accountability. **Even as Governors and Chief State School Officers were implementing reforms, they were becoming increasingly aware that they had no adequate means for assessing their effects.**

By 1985, both the Council of Chief State School Officers and the National Governors Association recognized the need for comparable data on education performance. Both organizations passed resolutions advocating the need for state-comparable data and the modification of the National Assessment into a source of such data. In 1987, a study panel on the National Assessment headed by then Governor Lamar Alexander and H. Thomas James issued a report to the Secretary of Education. The "single most important change" recommended by the commission was that it collect state-representative data. Another important recommendation was to change the governance structure of the National Assessment so that policy, administrative management, and testing operations were separated functionally and organizationally.

In 1988, the Congress reauthorized the National Assessment, providing for trial state assessments in 1990 and 1992 and fashioning a new governance structure. The new law established the National Assessment Governing Board (the Governing Board) to formulate policy, assigned administrative responsibility to the National Center for Education Statistics (NCES), and authorized the conduct of the National Assessment through grants, contracts or cooperative agreements.

The desire for checks and balances and for policy to be formulated in the "sunlight" appears evident from the National Assessment legislation. The new three-way structure replaced the previous approach in which a single organization received a grant and controlled all aspects of the assessment: setting policy, conducting the assessment and reporting results. Under the new structure, National Assessment reports are prepared by the contractor, but are subject to policies established by the Governing Board, and overall direction and quality control reviews performed by NCES. The inner workings of the National Assessment are public. For example, the Governing Board is required to conduct its business in public--all of its meetings, by law, are subject to federal open meeting requirements. Thus, policy formulation by the Governing Board, implementation and reporting plans of NCES, and the activities of the contractor all are matters of public record. In practice, the Governing Board invites public comment before acting on major policy matters.

Beginning in September, 1989 with the Education Summit convened by President Bush and the Governors, the context for education policy has changed even further. Major events include:

- o Announcement by the President in the 1990 State of the Union address of six national education goals, developed in concert with the Governors;
- o Creation in July 1990 of the National Education Goals Panel to monitor progress toward the goals;
- o Establishment in spring, 1991 of the National Council on Education Standards and Testing (NCEST) to examine the desirability and feasibility of developing and implementing national education standards and a system of assessments; and
- o Release in January 1992 of the report of NCEST, which recommended the establishment of national education standards, creation of a national system of assessments, and the use of the National Assessment to monitor national and state progress toward national education goals 3 and 4.

The assumptions and circumstances underlying education in the 1990s are in stark contrast to those in which the National Assessment was conceived almost three decades ago. Today there are national goals for education, the prospect of voluntary national standards for content in the subject areas and for student performance, and the possibility of system performance and school delivery standards. Where the idea of a national curriculum was anathema thirty years ago, and continues to be a concern for many, it is clear that the concern has lessened. Where there was great resistance to comparable data and reporting in the 1960s, it is a widely accepted idea now, at least at the state, national and international levels.

It is likely that these developments will have a bearing on the National Assessment. For example, in January 1992, the National Education Goals Panel resolved that, through this decade, the National Assessment should be the primary source of data for measuring national and state progress in student achievement in grades 4, 8 and 12. This refers in particular to progress in relation to Goal 3, which states, in part, that

By the year 2000, American students will leave grades four, eight and twelve having demonstrated competency in challenging subject matter...

Additionally, the advent of national education standards and their application in defining "competency" would necessarily have an effect on the content of the assessments.

Over the years, the National Assessment has kept pace with a number of developments: the incorporation of standards for mathematics developed by the National Council of Teachers of Mathematics, more performance items, faster release of data, and state-level reporting. Many of these changes reflect policies adopted by the Governing Board, and many of the issues raised in this paper have been addressed by the Board. But this is a time of transition, a time to take a long look that leaves no assumption about the National Assessment unexamined.

With this as background, the Governing Board is attempting to address several fundamental issues concerning the design and use of the National Assessment. The Governing Board's aim is to prompt evaluation, debate and full consideration of the validity and utility of the assumptions underlying the National Assessment.

Role and Purpose of the National Assessment

For more than twenty years, the National Assessment has been the only continuing, nationally representative assessment of education performance. Its purpose has been to describe what U.S. students know and can do in important areas of the curriculum. Generally, the content of its tests has been designed to reflect instruction students are likely to experience; its aim has not been to lead instruction. The National Assessment has served as an effective barometer of national performance and of trends in performance. Its results have documented improvements in minority student performance since the 1970's, but generally level performance overall.

By design, the National Assessment is a "low-stakes" test for those who participate. Its results are reported at national and regional levels; no student, school, or district results are reported. Links between National Assessment results and state and local programs and policy are generally weak and unsystematic. Teachers, administrators, policymakers and the public often do not find its results relevant to their immediate concerns--improving the performance of their students and schools.

Many view the role of the National Assessment as a strength, advocating its use as a dispassionate monitor of national student performance. They argue that, because it is divorced from direct instruction and unencumbered by accountability pressures, the National Assessment provides a reliable indicator of performance and trends. Others view its current role as an impediment to its utility, questioning its ability to

inform instructional improvements in the classroom and education policy. They believe that because National Assessment results are reported at the national and regional levels only (except for the trial state assessment, discussed below), far removed from the classroom and from policymaking, they have very limited applicability for teachers, school administrators and state policymakers.

State-level Reporting

Much of the debate on its role involves one central issue--whether the National Assessment should collect and report data at a level capable of informing instruction and policy. Specifically, whether the National Assessment should regularly report information for states and should permit districts and schools to use the National Assessment for reporting local results.

Prior to its reauthorization in 1988, states and school districts were permitted to augment the National Assessment sample at their option and cost to obtain state and local results. States and districts that used this option did so because of the quality, subject coverage and innovative methods of the National Assessment. Performance assessment, assessment in the arts, and matrix sampling were among the innovations pioneered by the National Assessment.

In 1988, the Congress authorized state assessments on a trial basis. This congressional action followed recommendations from the National Governors Association, the Council of Chief State School Officers and the Alexander/James Study Group, all endorsing the use of the National Assessment for reporting state-representative data. These recommendations were informed by the results of an initiative conducted during 1985-1987 by the Southern Regional Educational Board. This initiative used National Assessment reading and writing tests to measure and compare the performance of eight volunteer states.

The congressionally authorized trial provided for assessments in 1990 (eighth grade mathematics) and 1992 (fourth grade reading and fourth and eighth grade mathematics), voluntary participation by states, and an independent evaluation of the feasibility and validity of these assessments. However, the 1988 legislation also prohibited the use of the National Assessment for reporting district or school-level results. As before, individual student results also are prohibited from being reported.

An evaluation of the 1990 trial state assessment in eighth grade mathematics was prepared by a panel of the National Academy of Education (NAE). The NAE panel found that the trial was successful, but too limited in scope to support a conclusion at this time that state assessments should become a regular part of the

National Assessment. The NAE panel recommended expanding the trial in 1994 to mathematics, reading and possibly science at grades 4, 8 and 12. The Governing Board and the Administration endorse the expansion of the trial state assessment to three subjects/three grades in 1994. In addition, the Governing Board and the Administration have recommended that, once shown to be feasible and valid, state assessments should be made a regular option of the National Assessment.

There are other proponents of state-level reporting. The National Council on Education Standards and Testing recommended that the National Assessment be used to monitor the Nation's and states' progress toward Goals 3 and 4 of the National Education Goals. The National Education Goals Panel accepted this recommendation and has decided to use the National Assessment as a primary source of data for monitoring student achievement at the national and state level. However, the Congress has not authorized state-level reporting beyond the 1992 trial.

Many who support the use of the National Assessment for state-level reporting believe that its potential for informing instruction and policy is not sufficiently tapped. They believe that comparable state data can be useful in assessing the productivity of education systems, the impact of curricular policies, and the identification of strengths and weaknesses among various student subpopulations. States that participate could use the National Assessment to chart trends in student performance and as an indicator of the impact of state policies and programs. States that choose could use it to compare the performance of subpopulations in their state with the performance of similar subpopulations in other states. For example: Alabama could compare the performance of its female students with those in Kentucky or Maine; California could compare the performance of its Asian students with those in Illinois or Georgia; and New York could compare the performance of its disadvantaged urban students with those in Ohio or Michigan. The primary purpose would be to help states identify effective policies and programs.

However, there are arguments against using the National Assessment to report results at the state level. Some are concerned that this could lead to misuses of NAEP data, among which they would include: reporting state "rankings," meting out rewards or consequences (whether federal or state) on the basis of National Assessment results, and inappropriate high-stakes applications. Another often expressed concern is that state-level pressures to perform well on the National Assessment could lead to a narrowing of curriculum focused on NAEP test objectives, corruption of test results by teaching to the test, and a de facto national curriculum. Others are concerned that state-level testing and reporting by the National Assessment could have the effect of limiting state assessment research and development, which in recent years has helped

advance assessment methodology. Additional issues raised are that student populations among states are not necessarily comparable, that resources devoted to schools are not necessarily comparable, and that comparisons among schools, districts or states do not in themselves improve achievement. The policy debate on the role of the National Assessment must address directly whether the protections built into the National Assessment to prevent federal intrusion also have the effect of impairing its ability to accomplish Commissioner Keppel's vision that it "furnish a basis for public understanding of educational progress and problems." Because NAEP's effect on education policy and practice is limited, it cannot be said that this purpose has been fully accomplished.

Costs and Test Burden

Costs and testing burden are two additional concerns. For the 1990 and 1992 trial state assessments, the costs of sampling, test administration training and monitoring, data analysis, and reporting are paid by the federal government. However, the cost of test development (i.e. test framework and specifications, item development and field testing, and printing of test booklets), which is required by the nationwide administration of the National Assessment, is not an additional cost. The cost of test administration currently is borne as in-kind costs by the participating states.

Each assessment involves a multi-step process that takes about five years to complete. Thus, the cost of an assessment is distributed over a five-year period. The five-year cost of an assessment at the national level in three subjects at three grades is about \$18 million. The additional cost of conducting the National Assessment at the state level is approximately \$58 million, for a total of \$76 million. Although the costs are not distributed equally over the five years, on an annualized basis this would amount to about thirty-three cents for each U.S. school child for a three subject/three grade assessment at the national and state levels. The in-kind costs of administration have been estimated by various states participating in the 1990 trial state assessment, in eighth grade mathematics. The estimates range from about \$50,000 to almost \$200,000 per state. Although these estimates are not comparable, being based on different assumptions, they do indicate the costs of administration for one subject in one grade. With proposals to expand state assessments to three subjects and three grades in 1994, these costs could rise, although not necessarily proportionally. The Education Department has proposed legislation under which these costs would be shared--the federal government would pay a state's costs of administration, in excess of the first \$100,000. The Education Department has estimated the cost of supporting state administration in three subjects and three grades at about \$46 million. These costs would be spread over two years (split 20 percent in the pre-test year and 80 percent in the year the tests are given).

The size of the state sample is about 2,500 students per subject per grade. This size sample permits fairly accurate estimates of performance by subpopulations (e.g. gender, race, level of education attained by parent, etc.). However, in small states, such as Delaware, the sample-size requirement can result in testing all students in a grade. As subjects and grades involved in state-level assessment increase, many students could be tested more than once. In cases in which state assessment programs cover the same grades as the National Assessment, testing of students could become very problematic. The Council of Chief State School Officers (CCSSO) has recommended continuing state-level assessment and has suggested that states be given the option to select the components of an assessment in which they participate. Thus, Maryland might decide only to participate in one subject/one grade rather than multiple subjects in the three grades tested by the National Assessment. Another option would be to reduce the sample size in a small state; however, doing so would limit the ability to estimate the performance of subpopulations within a state.

Accountability

Finally, the examination of the role of the National Assessment must consider the question of its use for accountability purposes. Recent developments, including the establishment of national education goals and widening support for national education standards and a system of assessments, represent a departure from past assumptions about the American education system and raise fundamental questions about the role that the National Assessment can and should play. An important consideration is the National Education Goals Panel's decision to use the National Assessment as a primary source of data for monitoring student achievement at the national and state levels. Over time, through regular reporting and public attention, this use of the National Assessment is likely to have accountability effects; if this is to be the case, it should be accomplished as a matter of policy, by design and with purpose, not by happenstance.

Policy Questions:

Should the National Assessment continue as designed in the 1960s as a national and regional monitor only?

Should the National Assessment regularly collect and report state-level data?

Should the full costs of collecting and reporting state-level data be borne by the federal government, shared with the states, or paid by the states?

What options should be considered for reducing sampling burden on small states?

Alignment of the National Assessment with Nationally Certified Content and Student Performance Standards

The call for voluntary national content and performance standards poses several challenges for the National Assessment. The National Assessment has focused on describing what students actually know and can do, but the assumption is that national content and performance standards will express a vision of what students should know and be able to do. This assumption bears on the content of NAEP tests. Currently, the Governing Board conducts a national consensus process to develop the framework, objectives and specifications for each assessment. The consensus process is based on the assumption that curricula differ among the states and that a broad consensus on the content of each assessment is essential. This is particularly true in assuring that state-level reporting is valid and fair. Through the consensus process, a balance is achieved between current practice and advances in the discipline, based partly on research and development and on the views of experts.

The advent of national content and performance standards would pose a fundamental question: Should these standards determine the content of each assessment or should the National Assessment attempt to reflect these standards and the current (and evolving) instructional programs? Choosing the former could make the National Assessment a measure of implementation of the standards, but might provide less information about what students actually do know until the standards are widely implemented. In such a role, the National Assessment might be viewed as leading instruction and as an impetus for adopting the standards. Choosing the latter approach would more closely conform to its current role, in which current practice and aspirations for the disciplines are both incorporated in the National Assessment. This latter approach requires incremental adjustments in NAEP frameworks over successive assessments; it would not necessarily hasten state and local adoption of the standards, and could raise questions about inferences that properly can be made about progress toward the standards.

Adoption and implementation of national standards would be accomplished in different ways and at different rates throughout the U.S., especially since it is recommended that these standards be voluntary. If the Congress continues to authorize state-level reporting, the fact that standards are voluntary may have implications for the National Assessment. On the one hand, it is arguable that overly abrupt incorporation of new standards might discourage some states from participating, concerned that they were not far enough along in implementing the standards. On the other hand, states might welcome the National Assessment taking the initiative as a way to prompt local implementation.

Another dilemma relates to timing. It is likely that voluntary national content and performance standards would be developed according to timelines independent of those for the National Assessment. The cycle for an assessment now encompasses about five years from planning to reporting, with work to develop the frameworks beginning at least four years prior to an assessment. With no certainty that a set of standards would be developed and certified by a fixed date, the feasibility and desirability of significantly altering National Assessment plans in the absence of such standards could become a concern.

Work on education standards began in mathematics, is underway in science, history, geography, civics and the arts, and is expected to occur in English as well. Thus, how the National Assessment should handle new, voluntary standards as they are developed and adopted is an immediate issue. The current approach is based on the fact that standards are being developed and on the assumption that they will be subjected to a formal certification process that will give them standing in the education community. Clearly, the congressional mandate to develop the National Assessment through a national consensus process would appear to require attention to such standards in preparing an assessment. The Governing Board believes that appropriate alignment of the National Assessment with certified national standards is essential, that national standards should be a primary basis for developing assessments, that incorporation of such standards into the National Assessment should be done through successive adjustments of its frameworks and assessments, and that the goal should be to achieve a balance between the vision contained in new voluntary standards and the reality of current instruction.

The Governing Board has already taken this approach with the mathematics standards developed by the National Council of Teachers of Mathematics (NCTM). The framework for the 1990 mathematics assessment drew heavily upon the NCTM standards while they were in draft form. The 1992 mathematics assessment was revised, taking the final NCTM standards into account, with more than 50 percent new items. Specifications for the 1994 mathematics assessment have been developed that further incorporate the NCTM standards.

Although the Governing Board has followed what it believes to be a prudent course, issues remain to be resolved about the appropriate link between national standards and the National Assessment. The central question to be resolved is whether the National Assessment should reflect "what is" or "what should be." The Governing Board views this as an open issue at this time.

Policy Questions:

Should national content and performance standards determine the content of each assessment or should the National Assessment reflect these standards and the current (and evolving) instructional programs?

What formal mechanisms should the National Assessment institute in order to collaborate with organizations as they develop standards?

Assessment Frameworks

A related issue involves the frameworks that the Governing Board is responsible for developing for each assessment. Even if national standards do not come into being, there will still be tension between "best practice and aspirations" in a discipline on the one hand, and "predominant instructional practice" in the classroom on the other.

The Congress requires that the National Assessment frameworks be developed through a broadly representative national consensus process. The national consensus process has proven reliable and effective. It has resulted in assessment frameworks that are inclusive, that integrate competing views of curricular approaches, that represent high expectations, and that are recognized for their quality and innovation.

In recent years, the national consensus process has resulted in frameworks that have begun to evolve in a direction inclined more toward leading than reflecting education practice. This evolution has been influenced greatly by the adoption of national education goals and by the movement toward national education standards. But it is independent of those efforts, arising from the national consensus process itself. However, it is a truism that what is tested is attended to, and thus, the content that is covered in the National Assessment can be a matter of great significance.

One issue that arises concerns monitoring education progress. If National Assessment frameworks are designed in ways that lead instruction, the ability to track performance trends over time could be weakened by abrupt changes in the content and coverage of assessment items. As one prominent psychometrician put it "When measuring change, don't change the measure."³ This is not a new problem for the National Assessment and has a straightforward solution. Where a subject has been previously assessed and a new framework is warranted by developments in the

³ Beaton, Albert E., unpublished paper, 1992.

discipline, a two-fold approach is taken. First, a trend assessment is conducted, using test items from previous assessments. Second, a "cross-sectional" assessment is conducted, using new items based on the updated framework. Over time, assessments based on the new framework can be used to establish a new trend line.

Another issue concerns the subjects included in the National Assessment. Under current law, the National Assessment includes reading, mathematics, science, writing, geography and U.S. history. In addition, the Congress authorized the Governing Board to select other subjects for assessment. Since the selection of a subject for assessment could influence the relative emphasis it is given in instruction, the question of what subjects the National Assessment assesses is also important.

Policy Questions:

Should the National Assessment be designed in a way that leads instruction (as developed through a national consensus approach) or should it reflect representative instructional practice?

Should the National Assessment assess subjects in addition to reading, mathematics, science, writing, geography and U.S. history?

Should the approach to measuring trends be modified?

Role of the National Assessment in Relation to Organizations that May be Established to Review or Certify National Standards and a System of Assessments

The National Council on Education Standards and Testing (NCEST) recommended the creation of a coordinating body, appointed by the National Education Goals Panel, to ensure the establishment of national education standards and a system of assessments. This coordinating body would be known as the National Education Standards and Assessments Council (NESAC) and would establish guidelines for national education standards and for assessment development. The Goals Panel and NESAC would have joint responsibility for certifying national education standards and determining whether assessments were appropriately aligned with those standards. The Goals Panel is working with the Congress on this proposed coordinating council, but authorizing legislation for NESAC has not been enacted. The Goals Panel is preparing a slate of candidates from which appointments to an interim group could be made as early as the late summer 1992.

The role of the Governing Board, as mandated by the Congress, is straightforward--formulating policy for the National Assessment and taking appropriate actions to improve its form and use, among other responsibilities. The role of the Governing Board does not include setting content standards for the curriculum used in the nation's schools or setting performance standards for making judgments about individual students. The Congress mandated that the Governing Board perform its functions independent of the Secretary of Education and the Department of Education and that it "exercise its independent judgment free from inappropriate influences and special interests." The Governing Board is comprised of elected and appointed public officials, education policymakers, teachers, test and measurement experts, school administrators, representatives of business and industry, and the general public, including parents. The composition of the Governing Board, consistent with its charge, is broadly representative to assure against any tendency toward federal intrusion in local decision-making and against undue influence from any source.

The NCEST proposed, and the Goals Panel agrees, that the National Assessment should be used to monitor national and state progress toward National Education Goals 3 and 4 and should be aligned with national education standards as they are developed. These recommendations are embodied in H.R. 4323. H.R. 4323 makes NAEP's only purpose "measuring the Nation's progress in meeting the national education goals..." H.R. 4323 also gives the Governing Board the responsibility for "ensuring that [the National Assessment] is aligned with national content standards." Some of the implications of these recommendations have been discussed earlier in this paper. However, the role of the Goals Panel as an audience for National Assessment data, along with its proposed joint role with NESAC to review and/or certify voluntary national education standards and a national system of assessments, raise additional issues related to governance of the National Assessment.

The NCEST proposal suggests that the review/certification process will be voluntary. States, districts, and commercial testing companies would, at their discretion, submit education standards and assessments to a National Education Standards and Assessment Council. The Goals Panel, in using the National Assessment to monitor progress toward national education goals, would have a special interest in the alignment of the National Assessment with national standards.

Checks and Balances

The Governing Board has maintained that, in the development of national standards and assessments, checks and balances are needed to assure that the public interest is protected, that such standards and assessments are valid and implemented

with integrity, and that authority over education decisions is not inadvertently or inappropriately centralized. The Governing Board also has encouraged widespread use, evaluation and secondary analysis of National Assessment data. The underlying concern is the independence and integrity of the National Assessment, for which both checks and balances and evaluation are needed. What is as yet unclear is whether these values would be advanced or come into conflict as far as the review/certification of the National Assessment by the Goals Panel and NESAC would be concerned. That is, whether such an evaluation process would have the effect of promoting or impairing the independence and integrity of the National Assessment.

As a federal program, the National Assessment is accountable to the public. A part of that accountability involves the availability of test items, data and background materials to the public. This is both for purposes of disclosure and to facilitate secondary research. For each assessment, some NAEP items are released to the public, but a portion are not released in order to protect the reliability and validity of the assessment. However, these "non-release" items may be, and have been, released to individuals and organizations for specific purposes with the permission of the Commissioner of Education Statistics. Thus, because National Assessment test items are in the public domain and are a centerpiece for reporting progress toward the National Education Goals, it is not inconceivable that the Goals Panel and NESAC could take the initiative to review and certify the National Assessment.

On the one hand, it could be argued that the National Assessment should be subjected to a review/certification process as another check and balance to assure its appropriateness for measuring education performance, especially if it is to monitor national education goals. On the other hand, it could be argued that such a review/certification process has the potential for undermining the National Assessment.

For example, the review/certification process would be conducted according to timelines distinct from those for the National Assessment. Would the Commissioner of Education Statistics be obliged to defer the conduct of an assessment if the review/certification timeline did not coincide with the assessment timeline? Would the review/certification process supplant the national consensus process that is used to develop NAEP frameworks?

Obviously, these questions cannot be answered yet because no review/certification process has been established. However, it is clear that the establishment of national standards and review/certification procedures for assessments bears on how the National Assessment would be administered. Therefore, it is important that, with respect to the National Assessment, the reach of the

review/certification process envisioned by NCEST for the Goals Panel and the proposed NESAC be fully discussed and debated.

Policy Questions:

Should the National Assessment, through the Governing Board and the National Center for Education Statistics, be treated the same as other assessment programs by any new entity created to review/certify standards and assessments?

Should there be a relationship specified in law between the National Assessment and any new entity created to review/certify standards and assessments? If so, what should that relationship be?

Achievement Levels for the National Assessment

The 1988 National Assessment legislation assigned the Board the duty to "identify appropriate achievement goals for each age and grade in each subject area tested under the National Assessment." This was a new requirement. It was not mentioned in the legislative history, thus the language in the law was the only guidance or direction the Governing Board had about what the Congress intended.

The Governing Board recognized this requirement to be a departure from then traditional practice. A long-standing NAEP design principle had been to report what students know and can do; now the Governing Board understood that it was to develop the means for reporting results in terms of what students should know and be able to do.

In its first years, National Assessment reports listed each item and the percentage of students answering that item correctly. In the mid-eighties, the NAEP contractor, Educational Testing Service (ETS), developed some reporting innovations. ETS translated the raw scores into a single, cross-grade scale from 0-500 and reported average scale scores for various populations. The midpoint of the scale is about 250 and "anchor levels" are set at 50-point gradations above and below the mid-point (a standard deviation is equal to 50 scale points). However, in neither case--reporting percents correct or scale scores--is it possible to determine whether performance is "good enough."

For example, knowing that fewer than 50% of 17-year-olds know in which half-century Abraham Lincoln was president provides a fact, but no context for interpreting that fact. Likewise, reporting that from 1973 to 1990, 13-year-olds' scale scores in

mathematics rose from 264 to 270 tells little about the math content of either score, what math content is represented by the increase, and whether that math content is demanding. On the other hand, using achievement goals would result in reporting results in terms of what students should know, and thus improve understanding of National Assessment data.

After more than a year of study, consultation and public comment, including consideration of 38 different approaches, the Governing Board decided on a step-by-step approach to set what it termed "achievement levels."

One alternative the Governing Board considered was to determine benchmarks using NAEP "anchor levels" and then set improvement targets as "X"% more students reaching a particular level. The Governing Board rejected this approach for several reasons. First, there was no readily apparent, non-arbitrary way of determining a targeted percentage increase from assessment to assessment. Second, setting improvement targets would not fix the basic problem--that the anchor levels do not portray the quality and sufficiency of student performance, only the distribution of performance. Third, the Governing Board felt that setting improvement targets for the American education system was more properly a state and local responsibility, or perhaps a responsibility of the National Education Goals Panel; but it was not within the scope of duties assigned to it by the Congress.

On the basis of Governing Board discussions and public comment, the Governing Board decided to adopt three achievement levels for each grade and subject. Three levels rather than one were chosen so that results reported within grades would span the distribution of performance, not just provide a "pass-fail" mark. A tested standard-setting procedure, which employs a judgment process for setting "cut-scores" on a test, was adapted for establishing the achievement levels. The three achievement levels are: Basic (partial mastery); Proficient (competency over challenging subject matter); and Advanced (superior performance beyond Proficient). The proficient level is intended to reflect the reasoning used in framing National Education Goal 3 (which calls for students to demonstrate competency over challenging subject matter). National Assessment data are to be reported according to these levels, and states and the National Education Goals Panel could, if they choose, set their own improvement targets using these levels.

Achievement levels were developed first for reporting results of the 1990 mathematics assessment, as a trial, so that lessons learned could be applied to the Board's future efforts to set achievement levels. These achievement levels have been evaluated by the Governing Board, the National Academy of Education, the General

Accounting Office and others. The utility, reliability and validity of these achievement levels have been widely debated. The Governing Board will use the results from the 1990 mathematics achievement levels trial process as baseline data only if such use is found to be warranted; the Governing Board will consider the results of the 1992 mathematics achievement level-setting process in making this determination.

Informed by the evaluations of the 1990 trial process, the achievement level-setting process was improved. Achievement levels are being developed for the 1992 assessments in reading, writing and mathematics. The Governing Board believes that using achievement levels will promote public understanding of NAEP results and represents an improvement over previous reporting approaches. But the Governing Board is also aware that reporting what students should know and be able to do represents a significant shift in the National Assessment program. In addition, there is a division of opinion on the usefulness and appropriateness of achievement levels.

Policy questions:

Do achievement levels improve public understanding of the National Assessment?

Should the National Assessment results continue to be reported using achievement levels?

Should other approaches to identifying appropriate achievement goals be considered? If so, which approaches should be considered?

An International Component to the National Assessment

One of the underlying premises of the National Assessment is that it is to measure the performance of U.S. students in academic subjects and the change in their performance over time. However, there is growing concern about the achievement of U.S. students as compared to the performance of students in other nations. These include assessments conducted by the International Association for the Evaluation of Education Achievement (IEA) and by the International Assessment of Educational Progress (IAEP).⁴

⁴ Mathematics and science assessments were conducted by the IEA in the mid-1960's and 1982 and by the IAEP in 1988 and 1991. IEA will conduct mathematics and science assessments again in 1994 and 1998.

Some of these international assessments also have raised concerns about the quality of the curriculum and the level of expectations for U.S. students. Implicit in the call for national content and performance standards is the belief that such standards would help make the U.S. more competitive and that they should take into account the demands of the curricula of high performing nations and the performance of students in those nations. Thus, it is thought, the standards developed for the U.S. should be informed by the education practices and outcomes of other nations.

However, making valid international comparisons is a complex undertaking. Curriculum and instructional practices, language and culture, comparability and representativeness of student populations, and participation of disabled and disadvantaged students vary widely and prevent easy international comparisons. Currently, the National Assessment plans to perform equating studies so that its 1992 mathematics assessment results can be compared with the performance of other nations on the 1991 IAEP; similar studies are planned for comparing the 1994 National Assessment science and mathematics results with the 1994 IEA assessments. But international assessments are not conducted regularly and haven't covered the range of subjects taught in U.S. schools. Therefore, reliance on international assessments can only partially satisfy the interest in using international data to improve interpretation of U.S. education performance and to inform education policy.

There are several possibilities for the National Assessment. First, international comparisons of curriculum could be used in developing assessment frameworks. Similarly, international comparisons of student performance standards and test outcomes could be used in establishing or validating achievement levels set for the National Assessment. Finally, the performance of students in other countries on National Assessment tests or items, while not constituting an international assessment, could be useful in interpreting domestic results. This could be accomplished in a variety of ways with countries willing to participate, from administering National Assessment tests to comparable populations of students in other countries, to exchanging items with other countries, embedding the items in respective domestic tests, administering the tests under similar conditions and comparing results.

Policy questions:

Would it be useful to see how students in other nations perform on the National Assessment or on certain NAEP items as a reference point for understanding U.S. student performance?

Should the development of National Assessment frameworks regularly take into account the nature of curricula, instruction and expectations of foreign education systems?

Should the National Assessment achievement levels be developed taking into account student performance in other nations?

The National Assessment as an Anchor for Linking State and Local Assessment Systems with National and International Results

The report of the National Council on Education Standards and Testing recommends a two-component system of assessments: student assessments that provide results for individual students and large-scale assessments conducted on a sampling basis, such as the National Assessment. Further, the report recommends that the new system of assessments be aligned with national content and performance standards and have the capacity to provide comparable results (i.e. individual student, school, district, state and national). Finally, the report recommends that the National Assessment be reauthorized and assured funding to monitor state and national progress toward National Education Goals 3 and 4.

The report does not specify how variously developed, individual student assessments would be made comparable to each other and with the large-scale assessments. An assumption is that alignment with national content and performance standards would be sufficient to ensure comparability. While this is probably a necessary condition, it is not sufficient. For test results to be comparable, the test content must be comparable and the reliability of the tests must be comparable. It would not be appropriate to assert that two tests are comparable solely because they contain content drawn from a common set of standards. For example, it is possible to develop two items of differing difficulty that purport to be based on the same standard; if that is the case, the items would not be measuring the same things and could not truly be considered comparable. A similarly complex issue involves the fact that individual student assessments would be developed in decentralized fashion by various groups and organizations. By what technical procedures would this multitude of tests be proven comparable to each other? Finally, the National Council on Education Standards and Testing report does not address the issue of review/certification of assessments for comparability, whether an entity should be created for this purpose, and whether comparability should extend to the international level.

A number of approaches exist to link test results for comparability, varying in the technical rigor required and the inferences that can be drawn. The most rigorous form of linking is called "equating," a statistical procedure in which scores on two separate, but very similar, tests can be used interchangeably. A less rigorous form of linking is "prediction" in which a score on one test would predict performance on another, even somewhat dissimilar, test. These predicted scores are very dependent on the context (e.g. predictions for individual scores may not hold for group scores), and are not interchangeable. Non-statistical, or judgment methods can also be used, in which expert opinion is applied in making judgments about the comparability of student performance; such methods are especially appropriate in comparing performance tests, such as essays.

One approach to the comparability issue would be to equate the many locally developed tests to a single agreed upon "anchor" test. The anchor test would incorporate national content and performance standards; its test objectives and specifications would be known and its test items, under controlled conditions, could be made available. Local results could be equated to the anchor test.

With the National Assessment serving as the primary source of data for National Education Goal 3, with the possibility that trial state assessments will be continued and expanded in 1994, and with concrete plans already in place to link the National Assessment with international assessments conducted in 1991 and to be conducted in 1994 and 1998, it would appear that the National Assessment is one possible candidate for serving as an anchor test for purposes of comparability among locally developed assessments. Obviously, state-developed assessments or commercial tests could serve as anchors as well. The purpose of the discussion here is not to argue against consideration of other alternatives (and perhaps more than one alternative should be made consistently available), only to prompt discussion about whether the National Assessment should be one of the alternatives.

However, using the National Assessment as an anchor test raises questions about the conditions under which such use would occur. For example, it is a secure test; access to test items is tightly controlled. This prevents "teaching to the test items" and is one of the ways that its integrity is maintained. If the National Assessment were used as an anchor test, procedures would need to be adopted to protect its integrity. Since the National Assessment is a federal program, such procedures would most likely be overseen by a federal agency. It is unknown whether such procedures would encourage the use of the National Assessment as an anchor test (because of its quality and integrity) or discourage such use (because of costs, administrative burden and resistance to federal oversight); the likely preferences of

schools, districts and states need further exploration. However, one example is currently underway as a trial. The Kentucky Education Reform Act requires the development of an assessment system that provides NAEP or NAEP-like results. The Kentucky Commissioner of Education proposed to the Department of Education that an equating study be done so that Kentucky assessment results could be equated to the National Assessment. The equating study is being conducted under carefully controlled conditions set by the Commissioner of Education Statistics and agreed to by Kentucky.

Another question concerns the National Assessment matrix design. The matrix design is intended for reporting group, rather than individual results. If the National Assessment were used as a comparability anchor for individual student tests, procedures would need to be developed to assure that inferences made are valid, reliable and fair and that decisions about individual students (e.g. promotion, awards, etc.) are appropriate.

Policy questions:

Should states and local districts be permitted to use the National Assessment as an anchor test for comparability purposes?

Should the federal government provide resources to permit research and development for such uses of the National Assessment?

Should there be federal oversight of such uses of the National Assessment to protect the integrity of the National Assessment, to avoid abuses and to assure that linking procedures are properly conducted?

Removing the Prohibition against Using National Assessment Results at the District or School Level

In 1988, in connection with the trial state assessment, the Congress enacted a prohibition against the use of the National Assessment for reporting results at the individual student, school or district levels. Prior to 1988, many states and districts, at their option and cost, augmented the National Assessment sample to obtain local (but not individual student) results. States and districts "bought in" to the National Assessment because of its rich, useful and unique data-base, and because of its credibility, integrity and innovativeness. Many states and districts have indicated that they would make use of the National Assessment to report local results if the prohibition were lifted. Some argue that lifting the prohibition, by permitting results

at levels where key instructional and policy decisions are made, would make the National Assessment more useful for teachers, administrators and policymakers. Others believe that the low-stakes nature of the National Assessment, by not motivating students to put forward their best efforts, may underestimate student performance. They suggest that the motivating effects of providing results to schools and districts could make it a more valid assessment.

However, there are a number of arguments against removing the prohibition. Some believe that, if the prohibition were lifted, in time the National Assessment would become a national test. Further, if it became a national test, too much power or authority over local education decision-making could devolve to the administrative, policy and operational apparatus responsible for the National Assessment (i.e. the National Center for Education Statistics, the Governing Board and the NAEP contractor). Again, the argument is that this would be a further step toward a national curriculum.

Some believe that the integrity of the National Assessment necessarily would be eroded if its use changes from general indicator to direct measure of student performance, and if it had high-stakes consequences. Their concern is that instruction would be narrowed and that "pressures to raise scores" could lead to teaching to the test and cheating. Others make the distinction between teaching to the "objectives" of a test, defining its content in general terms, and presenting actual test items in practice sessions with students. They deplore the latter, but argue that it is only fair and sensible to disclose the test objectives so that students are not denied the opportunity to learn that which is being assessed. In addition, some are concerned that voluntary use of the National Assessment at the local level could have two distorting effects:

1. Because of the presumed motivating effects of local reporting of results, samples drawn from states using the National Assessment for local reporting might not be considered equivalent to samples drawn from other states, thus reducing its reliability.
2. Again, because of the presumption of higher scores where there is local reporting, it might not be appropriate to compare local against national (low stakes) results.

However, others argue that statistical procedures could be used to identify and adjust for possible distorting effects. Such procedures were employed in reporting the 1990 trial state assessment data, because it had been found that state results were somewhat higher than national results.

Finally, some believe that administering it at the local level would increase data burden and turn the National Assessment into something unworkably large. However, based on the pre-1988 initiatives using NAEP to report local results, there is no reported evidence that this was a major problem (nor for that matter, that NAEP's integrity was eroded). Others argue that, as a voluntary program conducted at local cost, states and districts will consider "data burden" in deciding whether to pursue this option. It is true that the National Assessment is an extremely complex assessment program. The National Assessment uses matrix sampling, includes difficult-to-score performance items, and employs very sophisticated statistical procedures, all of which increase costs and time. However, some suggest that options for efficiencies exist. For example, instead of waiting up to a year or more for results (as is the case now for the National Assessment), districts could receive their results fairly quickly using a specially developed non-NAEP scale. This would provide timely district-comparable results. Later, when the National Assessment results are reported, district results could be "mapped" onto the NAEP scale.

Policy question:

Should states and local districts be allowed to use the National Assessment for local assessment at the district and school levels if they wish to do so under NAEP regulations and at their own option and cost?

Annual Assessments and Reporting

When the National Assessment began, it was an annual assessment. Assessments were conducted in a variety of subjects from 1969 through 1980. Beginning with the 1982 assessments, the authorizing legislation permitted assessments no more frequently than biennially. The National Education Goals Panel, the Administration, and the Governing Board have recommended changes to the authorizing legislation that would permit NAEP to conduct assessments in some subjects each year. For example, assessments in three subjects per year would permit a triennial cycle during which the six mandated subjects under the National Assessment (reading, writing, mathematics, science, history and geography) and three non-mandated subjects (e.g. civics, the arts, world history) could each be assessed once. The Council of Chief State School Officers (CCSSO) has recommended state-level reporting of two subjects on an annual basis to distribute data collection activities over two years, thus reducing the intensity of testing. However, CCSSO also urges that, with regard to state-level assessments, states be given the right to select the portions of an assessment in which they will participate (e.g. instead of participating in all subjects and grades in an assessment year, a state might want to have state data only

in one subject and grade). From the perspective of the National Education Goals Panel, annual assessments are needed to provide the public with regular, predictable reports on progress toward National Education Goal 3. Additionally, the Department of Education and the Governing Board believe that there are significant efficiencies that would accompany annual assessments.

Biennial assessments reduce efficiency. State participation in the National Assessment, being voluntary, is not subject to regularized operating procedures. Even though the same states tend to participate, they do not have on-going internal procedures, since National Assessment activities occur only once every two years. Thus, much more effort by the National Assessment operations contractor is expended in coordination than would be required on an annual cycle. Additionally, coordination within states has to be "re-invented" by state agencies every two years. For the National Assessment operations contractor, annual assessments would permit greater efficiency in deploying personnel than the current "hurry-up"/"slow-down" biennial cycle requires. Other efficiencies are possible with an annual cycle, including field-testing of new assessments and faster reporting. Such efficiencies could help keep down costs of the National Assessment program. In the long run, the total cost of a subject area assessment would not increase under an annual cycle of assessments, and might decrease.

Student testing burden could be an impediment to annual NAEP data collections. As mentioned previously, small states face the concern that too many students would be tested too frequently. In addition, more frequent National Assessment testing could foster logistical problems with state assessment programs. To the extent that the National Assessment competes with, rather than complements, state assessment efforts, state participation rates (and thus NAEP's robustness) could be affected. However, as noted above, CCSSO believes that annual assessments that give states choices in the degree of participation in a state-level assessment could reduce the testing burden.

Policy question:

Should the National Assessment legislation be amended to permit annual assessments?

CONCLUSION

The context of education policymaking has changed dramatically since the National Assessment's beginnings. The change is evident in the widespread interest in state data on education outcomes, the advent of national education goals and annual progress reporting, and the prospect of national education standards and a system of assessments.

The implications for the National Assessment--bearing on fundamental questions of its purpose and utility--could affect its underlying philosophy, design, and application. A report by the National Assessment technical advisory panel to the National Education Goals Panel recommended that the public should "...become at least as aware of National Assessment reports as they are of the Dow Jones Index or of the current unemployment figures." If that were to occur, the character of the National Assessment as an unobtrusive global monitor of education progress would change. The truth is that, despite more than twenty years of education reporting that has earned high respect among knowledgeable individuals, the vast majority of teachers, principals, and the public are unaware of the National Assessment. One can speculate about the reasons, but it is clear that the design of the National Assessment has minimized its relevance to local policy and instruction.

The National Assessment has responded to, or been influenced by, some of the changes in its midst. Grade-level reporting has begun and may in time replace reporting by age. The first trial state assessment was found to be feasible and to produce valid data; there is strong sentiment among users to continue state-level reporting. NAEP frameworks are widely acknowledged as responsible, representative and forward-looking. The Governing Board has received thousands of requests for its reading, writing, mathematics, and science frameworks and it is likely that NAEP frameworks will be used to inform the development of national standards. The use of the National Assessment, by the National Education Goals Panel, for monitoring achievement will heighten public attention to NAEP data and the pressure on the National Assessment to serve as an accountability mechanism, at least at the state level.

Should the National Assessment be used for accountability purposes at any level of reporting? Should NAEP test frameworks (as opposed to test items) influence classroom instruction? Should the National Assessment continue as an unobtrusive measure of achievement? Should the National Assessment be used in a way that promotes education reform and improvement? Answers to questions such as these will affect how the National Assessment is designed and how it is to be used in the future.

As noted above, there have been some changes in the National Assessment, although incremental in nature. However, even these changes challenge NAEP's original premises.

Issues that could affect the role of the National Assessment have been discussed in this paper. Although they have been discussed separately, many of the issues are interdependent. A decision on one issue could affect or narrow choices on another issue.

Two very broad options for the future of the National Assessment may well frame the debate. In one, the National Assessment remains in its current form as an unobtrusive national monitor, providing reliable information on education performance and progress. In the other, the National Assessment is recast as a component of a national assessment system, providing national data on student performance, permitting states and schools to participate in the National Assessment and receive results, and providing local and state education agencies with a mechanism for comparing the results of locally-developed assessments. Clearly, many permutations are possible and should be examined. However, to the extent that our nation is on the verge of establishing a system involving national education standards and related assessments, NAEP's appropriate place within that system deserves thoughtful consideration and public debate. The National Assessment Governing Board hopes that this paper will help prompt and inform the ensuing discussion.

EPILOGUE

The Governing Board intends to use comments it receives in developing its positions on these issues. During the late summer and early fall 1992, members and staff of the Governing Board will attempt to provide opportunities for interested organizations and individuals to express their views in writing and in person. These views will be analyzed and reported to the Governing Board. Currently, the Governing Board plans to consider adoption of positions on these issues at its meeting scheduled for November 19-21, 1992. Recognizing the importance of these issues for NAEP policy, and that the Congress will be considering legislative reauthorization of the National Assessment beginning in 1993, the Governing Board intends to provide a report of its activities related to the future of the National Assessment to the Administration and the Congress.

References

Alexander, Lamar and James, H. Thomas. The Nation's Report Card: Improving the Assessment of Student Achievement. Washington, D.C.: National Academy of Education, 1987.

Beaton, Albert E. Unpublished paper, 1992.

Hazlett, James A. A History of the National Assessment of Educational Progress, 1963-1973. Unpublished Dissertation, University of Kansas, 1973.

National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment. Assessing Student Achievement in the States. Stanford, CA, 1992.

National Commission on Excellence in Education. A Nation at Risk. Washington, D.C.: U.S. Department of Education, 1983.

National Council on Education Standards and Testing. Raising Standards for American Education. Washington, D.C., 1992.

U.S. Congress, Office of Technology Assessment, Testing in American Schools: Asking the Right Questions, OTA-SET-519 (Washington, D.C.: U.S. Government Printing Office, 1992).

Wirtz, Willard and Lapointe, Archie. Measuring the Quality of Education: A Report on Assessing Educational Progress (ERIC Document Number ED-213769). Washington, D.C., 1982.